

L 121937

Journal of Internal Medicine 2000; 248: 447–454

## INTERNAL MEDICINE IN THE 21ST CENTURY

### Molecular epidemiology

P. BOFFETTA

*From the Unit of Environmental Cancer Epidemiology, International Agency for Research on Cancer, Lyon, France*

**Abstract.** Boffetta P (Unit of Environmental Cancer Epidemiology, International Agency for Research on Cancer, Lyon, France). Molecular epidemiology (Internal Medicine in the 21st Century). *J Intern Med* 2000; 248: 447–454.

The use of biomarkers in epidemiology is not new, but recent developments in molecular biology and genetics have increased the opportunities for their use. However, epidemiological studies based on biomarkers, which belong to the discipline defined as 'molecular epidemiology' are subject to the same

problems of design and analysis as 'traditional' epidemiological studies. If biomarkers offer new opportunities to overcome some of the limitations of epidemiology, their added value over traditional approaches should be systematically assessed. Biomarkers should be validated through transitional studies; consideration to sources of bias and confounding in molecular epidemiology studies should be no less stringent than in traditional studies.

**Keywords:** bias, biomarkers, confounding, molecular epidemiology, statistical power.

### Introduction

The definition of molecular epidemiology applies to a discipline overlapping with both public health and experimental science. There is no universally accepted definition of which type of research fits into the discipline, and emphasis is given to different aspects that depend to a large extent on the background of the investigators involved. So, for an epidemiologist, molecular epidemiology might include any epidemiological study involving the use of any biologically based measurement (including possibly measurement of blood pressure), whilst for a molecular biologist the search for a new gene in a series of a few dozen patients would qualify as a molecular epidemiological study. This confusion may originate from the growing availability of molecular (and more in general biological) approaches to measure variables that might be relevant in epidemiology, and from the recognition

of the need to validate and apply them to carefully defined populations.

Despite the ambiguities in the definition of molecular epidemiology [1], the definition has been successful and is now widely used in names of departments, university courses and programmes, scientific meetings and professional societies. Several textbooks have also been published [2, 3].

The use of biologically based approaches to measure variables of interest in epidemiological studies is not new. Two areas in which studies have been conducted for decades that would nowadays be included under the rubric of molecular epidemiology are epidemiology of infectious diseases and cardiovascular epidemiology. As an example, Fig. 1 shows the results of one of the early study linking elevated serum cholesterol levels – which today might be classified as a biomarker of exposure – to risk of ischaemic heart disease in the Framingham cohort [4]. The recent expansion of studies based on

and the precision of the measurement of the biologically relevant exposure variables. In some cases, the biomarker-based measure of exposure represents an obvious improvement towards a better assessment of exposure.

Aflatoxin represents a good example in which exposure biomarkers have represented a step forward in the identification of human cancer hazards. The fungus *Aspergillus flavus* is a common contaminant of foodstuffs, in particular cereals and nuts, especially in West Africa and East Asia. In certain storage conditions, it produces a toxin, called aflatoxin, which shows strong hepatotoxic and carcinogenic properties in several animal models. Given the lack of evident colour, taste or smell of aflatoxin in processed food, there is no sensible way individuals can know whether the food they consume is contaminated. Studies on the carcinogenic effect of aflatoxin have therefore been limited by the difficulty in determining exposure status at the individual level, although ecological study areas (e.g. villages) in which contamination was frequent had a higher occurrence of liver cancer than neighbouring areas with less frequent contamination. This situation changed with the identification of serum and urine biomarkers of aflatoxin exposure, namely urinary metabolites of aflatoxin itself and of its adducts formed with DNA. Table 2 reports the results of the first investigation that assessed the risk of liver cancer in subjects with samples collected and stored before the disease occurred. Individuals with any urinary marker of exposure had a 2.4-fold increased risk of liver cancer relative to individuals without markers; the relative risk was as high as 4.9 amongst individuals positive for the urinary adduct degradation product AFB<sub>1</sub>-N<sup>7</sup> guanine [8]. It is noteworthy that urinary samples in this study were taken on average only two years before analysis, at a time that might not be biologically relevant for liver cancer development. It is therefore conceivable

that aflatoxin exposure might have been misclassified as compared with the biologically relevant exposure. Yet, the study provided strong evidence of a causal association between aflatoxin and liver cancer in exposed humans.

### Bias

Three main types of bias are recognized in epidemiology, and all three may operate in biomarker-based studies [9]. Selection bias arises from lack of comparability of groups included in the study (e.g. cases and controls). For example, exposed cases might be more (or less) likely to participate in a study than exposed controls. Information bias involves misclassification of participants with respect to disease or exposure status. In biomarker-based studies, information bias encompasses the issues of validity, reproducibility and stability of markers. Finally, confounding is a special form of bias, due to exposure to risk factors other than those under study (see below).

Selection bias can be avoided by properly identifying the study population, and by optimizing the response rate. Furthermore, it can be controlled in the analysis by identifying factors that are related to selection and by controlling them as confounders. Unfortunately, many molecular epidemiological studies pay relatively little attention to the selection of participants. This is particularly the case for studies of genetic factors, such as metabolic polymorphisms, because it is considered that any selection of participants is unlikely to be related to the genetic factors under study. As an example, an association with lung cancer risk has been reported in early studies of polymorphism of the CYP2D6 gene. Later studies, however, have not confirmed the finding, and the early results are likely to have arisen from the use of improper control groups [10]. In general, prospective studies, for example of the cohort design, are less prone to selection bias than retrospective studies, such as those based on a case-control comparison.

Sources of variation in biomarker-based measurements might arise from intergroup (e.g. cases versus controls) variability: this is the phenomenon molecular epidemiological studies usually aim to address. However, other sources of variation exist that generate misclassification. Interindividual variability might be due to genetic or environmental factors

Table 2 Relative risk of liver cancer and exposure to aflatoxin [8]

Biomarker	Cs	Co	RR	95% CI
No AF biomarker	9	87	1.0	Ref.
Any AF biomarker	13	53	2.4	1.0-5.9
AFB <sub>1</sub> -N <sup>7</sup> -guanine	6	11	4.9	1.5-16

AF, aflatoxin; Cs, cases; Co, controls; RR, relative risk; CI, confidence interval.

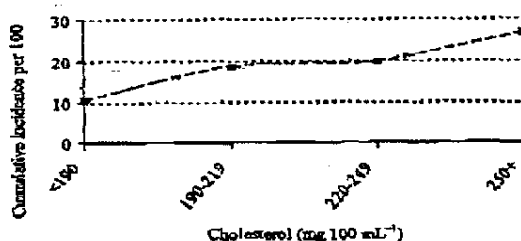


Fig. 1 Incidence of ischaemic heart disease and serum cholesterol in the Framingham cohort – 12-year follow-up of men [4].

biomarkers has represented a change in scientific paradigm in other areas of chronic disease epidemiology, notably in cancer epidemiology.

It is useful to consider molecular epidemiology within the framework of epidemiological studies in general. Epidemiology aims to identify determinants of disease (either risk factors or protective factors) and to quantify their role. This is done whilst taking into account (to the extent which is feasible) sources of random and systematic error (bias and confounding), as well as factors that modify the effect of the determinant of interest (Fig. 2). As an example, epidemiological studies have shown an increased risk of cancer of the oral cavity amongst alcohol drinkers [5]. This relation has been statistically significant (i.e. random error has been excluded) in several large studies, and it has been confirmed that no matter how alcohol drinking and oral cancer were measured (i.e. information bias originating from misclassification of exposure and outcome has been excluded) and in different types of epidemiolo-

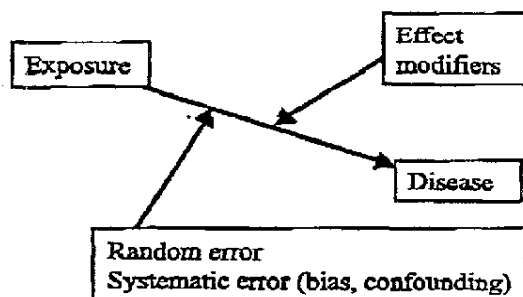


Fig. 2 Identification of exposure-disease relations with epidemiology.

gical studies (i.e. case-control and cohort studies, thus reducing the likelihood of bias from selection of study subjects). Furthermore, the potential confounding effect of tobacco smoking has been controlled in various studies (e.g. by restricting the analysis to nonsmokers), and the susceptibility to alcohol-related oral cancer in different groups (e.g. possibly related to ethnicity) has also been studied.

To a large extent, molecular epidemiological studies fit in the same framework: in other words, they consist of epidemiological studies in which either risk factors, outcomes, confounders or effect modifiers are measured with biomarkers. Similarly, the same arguments should be applied to judging the design, analysis and interpretation of these studies as in the case of 'traditional' epidemiological studies. One exception is the class of 'transitional' studies, which represent a type of investigation specific to molecular epidemiology (see below).

Biomarkers have been traditionally distinguished in markers of exposure, disease and susceptibility (Table 1). This distinction is, however, somewhat arbitrary. For example, chromosomal aberrations have been used for decades to monitor exposure to environmental carcinogens [6]. From this point of view, they can be classified as exposure biomarkers. However, recent evidence points towards a role of chromosomal aberrations in predicting cancer risk [7] in that subjects with an increased frequency of cells with aberrations are at increased risk of cancer, independently from the agent they were exposed to. In this respect, they can be seen as early markers of disease.

### Measuring with biomarkers

The rationale for using biomarkers to measure exposure lies in the attempt to increase the validity

Table 1 Classes of biomarkers

Class	Examples in cancer epidemiology
Exposure	DNA or protein adduct Viral DNA Urinary metabolite
Disease	Gene mutation Chromosomal aberration Precancerous condition
Susceptibility	Metabolic polymorphism DNA repair activity

Table 3 Sources of variation for selected biomarkers used in cancer epidemiology (adapted from Vineis, 1997 [11])

Biomarker	Intra-individual	Sampling	Laboratory
Hormones	+	-	(-)
Organochlorine compounds	-	-	+
DNA adducts in white blood cells	+	-	+
Chromosomal aberrations	?	?	+
Cervical dysplasia	+	+	+
Metabolic polymorphism (genotyping)	-	-	?
Metabolic polymorphism (phenotyping)	+	-	+

+, more important source of variation.

-, less important source of variation.

?, no or few data.

interacting with the variable under study. Intraindividual variability refers to components of variation such as daily variation in hormonal level. Finally measurement error might arise from sampling and laboratory variation. Table 3 provides some examples of sources of variation for selected biomarkers used in molecular cancer epidemiology.

Proper precautions should be taken to minimize the sources of variation other than intergroup variability. The potential sources of such bias are numerous: the circumstances in which biological samples are taken, processed, stored and analysed; the technical aspects of the assays, etc. It is important to ensure that, if all sources of variation cannot be controlled (as it is often the case), they should apply equally to the groups being compared. Therefore, if long-term storage of samples might affect the measurement, it is important to match cases and controls in the study by duration of sample storage. In such a case, misclassification is said to be 'nondifferential' (i.e. acting equally on the groups being compared). Non-differential misclassification invariably produces bias towards the null value, that is, it obscures an existing causal (or protective) association, but it does not generate false-positive results. On the other hand, a misclassification that is 'differential' with respect to case-control (or exposed-unexposed) status generates a bias in an unpredictable direction. For example, if there is substantial interbatch (or interreader) variability in the measurement, the inclusion of samples of cases and controls in different batches would generate differential misclassification, whilst a proper mix of samples in each batch would at most result in nondifferential misclassification

(e.g. because results of samples from one batch tend to be systematically different from those of samples from another batch).

### Transitional studies

The issue of variation in biomarker-based measurements impinges on the need to validate biomarkers before their application in large-scale studies. This is the domain of so-called transitional studies, which aim to characterize the biomarker itself rather than the underlying biological phenomenon. The aspects assessed by transitional studies include intra- and intersubject variability; feasibility of application of a biomarker in field conditions (and optimization of its use); confounders and effect modifiers for the marker; and underlying biological mechanisms reflected by the marker.

Transitional studies usually involve healthy individuals, patients or subjects with specific exposures (e.g. groups of workers). Three types of transitional studies have been described in the continuum between development of a new assay and its application in human populations (Table 4). Table 5 presents the results of an interlaboratory comparison of measurements of DNA adducts using the <sup>32</sup>P-postlabelling technique in blood samples from workers exposed to polycyclic aromatic hydrocarbons (PAHs) in foundries and unexposed subjects. The results suggest an important interlaboratory variability in this assay; lack of control for laboratory (i.e. a comparison of results amongst those exposed from laboratory 1 and amongst those not exposed from laboratory 2, or vice versa) would result in grossly biased results. Unfortunately, biomarkers are often

## INTERNAL MEDICINE IN THE 21ST CENTURY 451

Table 4 Types of transitional studies (adapted from Schulte &amp; Pearce, 1997 [12])

Type of study	Aims	Characteristics
Developmental	Development of biomarkers	Builds on experimental studies Test assay in human samples Evaluate biological sample collection, processing, storage Evaluate assay accuracy, precision
Characterization	Validation	Assessment of biomarker range in representative human populations Evaluation of external (or internal) exposure-biomarker relationship, biomarker kinetics, and potential confounders and effect modifiers
Applied	Use in cross-sectional, metabolic, panel studies	Evaluation of exposure status of various populations and further validation of the biomarker

applied in the field without a proper characterization and validation, and therefore hamper the interpretation of the results.

### Confounding

Confounding refers to a condition in which an observed association between a suspected risk factor and a disease is due to a different risk factor, which is a true cause of the disease. In a classical example, an association between tobacco smoking and cancer of the uterine cervix has been observed in many different populations. However, this association is likely to be confounded by infection with the human papilloma virus (HPV), which is a cause of cervical cancer and is associated with tobacco smoking (in the sense that in many populations smokers are more frequently positive for HPV than nonsmokers). The use of biomarkers does not prevent confounding from occurring, and it is important to consider confounding as an alternative explanation when associations are observed. In the example above, infection with HPV would be a confounder of the association between tobacco smoking and cervical cancer no matter how smoking, infection and cervical cancer are assessed (via questionnaires,

medical records, biochemical methods or molecular techniques). Furthermore, use of biomarkers might introduce confounders. Figure 3 presents the example of a study of occupational exposure to PAHs and lung cancer. Tobacco smoking might represent an alternative source of PAHs, of greater importance than occupational exposure. In such a case, the results of the biomarker-based test will be driven by tobacco smoking rather than occupational exposure, even in the absence of an association between smoking and occupational exposure.

### Interaction

Biomarkers have been widely applied to study gene-environment and gene-gene interactions in the pathogenesis of cancer and other chronic diseases. In general, an interaction between a genetic and an environmental factor can be studied using a 4-fold table as shown in Fig. 4. Individuals with the low-risk genetic trait and without the environmental exposure form the reference group, and the relative (or excess) risk is estimated in individuals with the high-risk gene, with the environmental exposure

Table 5 Validation study of DNA adducts in foundry workers and controls ( $^{32}$ P-postlabelling) – comparison of two laboratories [13]

	Laboratory	
	1	2
Exposed ( $n = 35$ )	$26 \pm 43$	$9.2 \pm 23$
Unexposed ( $n = 6$ )	$3.1 \pm 1.7$	$1.7 \pm 0.7$

Adducts are expressed per million nucleotides.

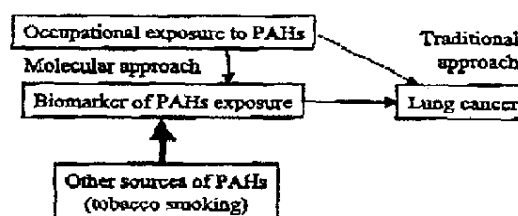


Fig. 3 Example of confounding.

	Low-risk gene	High-risk gene
Unexposed	1 (Ref)	RR <sub>g</sub>
Exposed	RR <sub>e</sub>	RR <sub>eg</sub>

Fig. 4 Interaction between an environmental factor (e) and a genetic factor (g). Ref, reference category; RR, relative risk.

and with both factors. Table 6 provides an example of a study of lung cancer that addressed both tobacco smoking and genetic polymorphism for the gene encoding for the enzyme glutathione-S-transferase (GST)M1, which might be implicated in the metabolism of tobacco carcinogens. The relative risk of lung cancer is higher in heavy smokers with the null genotype than in individuals with only one risk factor, suggesting an independent role of both tobacco smoking and GSTM1 polymorphism. In particular, the combined relative risk (10.2) is intermediate between what would be expected according to an additive model of interaction (assuming that tobacco smoking and the polymorphism act on different carcinogenic pathways, or  $RR_{eg} = RR_e + RR_g - 1 = 2.5 + 7.8 - 1 = 9.3$ ) and a multiplicative model (assuming that they act on the same pathway, or  $RR_{eg} = RR_e \times RR_g = 2.5 \times 7.8 = 19.5$ ). It should be noted, however, that the wide 95% confidence interval of the relative risk in the group with both factors (4.4–23.3) is compatible with both interac-

tion models, which stresses another methodological aspect of molecular epidemiological studies, namely the need for a large sample size (see below).

Molecular epidemiology studies addressing other types of interaction between two or more factors can be discussed according to the same paradigm used for gene-environment interactions. For example, in the study of aflatoxin exposure and liver cancer mentioned above, the investigators addressed the possible interaction of aflatoxin with hepatitis B virus (HBV). When compared to HBV-negative subjects with aflatoxin exposure, the relative risk in HBV-positive subjects who were also positive for aflatoxin markers was 60, which was greater than the product of the relative risks for the two factors separately (4.8 for HBV and 1.9 for aflatoxin), suggesting a synergism between aflatoxin and HBV in liver carcinogenesis. However, the wide confidence interval in the group with both exposures (6.4–560, based on only seven cases and two controls) does not allow the rejection of the hypothesis of no interaction according to a multiplicative model ( $4.8 \times 1.9 = 9.1$ ).

### Random error

From several of the examples quoted above it is clear that a major problem in biomarker-based epidemiological research is the insufficient number of subjects included in each study. The main reasons for a small study lie with logistical and financial constraints. Indeed, any biomarker-based measure introduced in epidemiology should be compared with traditional approaches (e.g. the assessment of a given exposure using a biochemical or molecular method versus a questionnaire), and the possible gain in sensitivity and specificity of the measurement should be considered in the light of the possible decrease in the number of study subjects.

Many authors have proposed formulae to calculate the sample size needed to detect main effects and interactions amongst risk factors [9, 15]. However, most published studies do not include enough individuals, resulting in unstable and often conflicting results. Recently, large-scale studies have started to be conducted (see, for example, a recent study of colon cancer and polymorphism for GSTM1 and NAT2 genes, based on over 4000 cases and controls [16]). Another approach is the pooling of indepen-

Table 6 Interaction between tobacco smoking and glutathione-S-transferase (GST) M1 polymorphism in lung cancer (adapted from Nakachi *et al.* 1993 [14])

Tobacco smoking	GSTM1 polymorphism	
	Wild type	Null
<44 pack-years	9/64	22/63
	1.0	2.5
	Reference group	1.1–3.7
44+ pack-years	24/22	30/21
	7.8	10.2
	3.3–18.2	4.4–23.3

In each cell, the first row shows the number of cases and controls, the second row the relative risk and the third row the 95% confidence interval.

dently conducted studies, as has been performed for studies of metabolic polymorphism and cancer [17].

### Publication bias

One characteristic of molecular epidemiology studies is the relatively large number of variables on exposures, disease and effect modifiers. Under these circumstances, the probability of generating by chance statistically significant results increases. It has been shown that there is a tendency to selectively report significant results, in particular when they show an effect in the expected direction. The net result is a biased reporting of positive over negative or null results. As an example, several studies have been conducted on polymorphism of the CYP2D6 gene, which encodes for an enzyme possibly involved in the activation of lung carcinogens and lung cancer risk. Figure 5 shows the results of the 16 studies available for a recent meta-analysis [18]. Results are reported in terms of logarithm of the relative risk for high-risk CYP2D6 polymorphism and of its standard error. Each study is identified by one dot; studies towards the right are smaller than studies toward the left, and studies towards the top are more positive than studies towards the bottom. If no publication bias exists, the pattern of such results should resemble a triangle (or a funnel), with larger studies converging on the left side around the central ('true') value, and smaller studies symmetrically dispersed on the right side. However, the empty side on the bottom right corner of the graph suggests that smaller studies were more likely to be reported if they showed a positive effect.

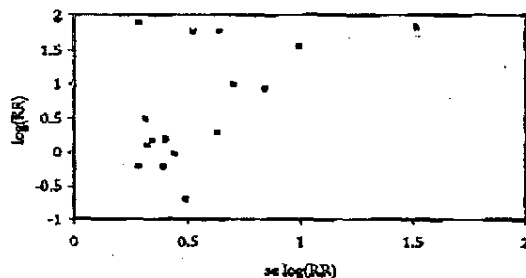


Fig. 5 Funnel plot of studies on CYP2D6 polymorphism and lung cancer.

© 2000 Blackwell Science Ltd *Journal of Internal Medicine* 248: 447–454

A formal test confirms the presence of an asymmetrical distribution of results.

It can be argued that such an initial report of false-positive results should not be considered a major scientific problem, since subsequent studies, aimed to replicate the early positive results, will eventually establish the truth. However, this approach is inefficient and represents an important waste of resources in particular in the case of studies based on expensive approaches, such as molecular epidemiological studies. For example, a study of metabolic susceptibility reported that postmenopausal, smoking women with slow acetylation genotype for the N-acetyl-transferase 2 gene had an increased risk of breast cancer, whilst this effect was not seen in nonsmoking women or in women with rapid acetylation genotype [19]. Given the absence of an overall increased risk of breast cancer from tobacco smoking, the results were not very plausible. However, many subsequent studies were published on this topic that failed to confirm the association.

A preferable approach consists of critically evaluating and reporting results on the basis of criteria other than (or including but not limited to) statistical significance. Biological plausibility, possible sources of bias and confounding, and number of tested associations are amongst such other criteria. Recently, statistical approaches have been proposed to take into account the possibility that significant results are generated by chance when many comparisons are made [20].

### Conclusions

Almost 20 years have passed since the term 'molecular epidemiology' was proposed [21]. It is clear that molecular techniques have found an important (and growing) role in epidemiological studies. So far, however, there are not many cases in which the application of a molecularly or even a general biologically based approach has represented an enormous leap in the evidence brought by traditional epidemiological methods. Assessment of exposure to aflatoxins, enhanced sensitivity and specificity of assessment of past viral infection, detection of protein and DNA adducts in workers exposed to reactive chemicals (such as ethylene oxide) are amongst the examples in which molecular epidemiology has greatly contributed to the understanding of human cancer. In many other

cases, however, initial, promising results have not been confirmed by subsequent, usually methodologically more sound, investigations. They include, in particular, the search for susceptibility to environmental carcinogens by looking at polymorphism for metabolic enzymes [22].

If biomarkers offer new opportunities to overcome some of the limitations of epidemiology, their added value over traditional approaches should be systematically assessed. Biomarkers should be validated; consideration of sources of bias and confounding in molecular epidemiology studies should be no less stringent than in other types of epidemiological studies. Similarly, other aspects of the study (e.g. determination of required sample size, statistical analysis, reporting and interpretation of results) should be approached with the same rigor as that used in epidemiology in general. When molecular epidemiological studies will be conducted according to state-of-the-art design, analysis and interpretation, the discipline will have reached its maturity.

## References

- Rothman N. Epilogue. In: Schulte PA, Perera FP, eds. *Molecular Epidemiology: Principles and Practices*. San Diego: Academic Press, 1993: pp. 565-9.
- Schulte PA, Perera FP, eds. *Molecular Epidemiology: Principles and Practices*. San Diego: Academic Press, 1993.
- Toniolo P, Boffetta P, Shuker DEG, Rothman N, Hulka B, Pearce N, eds. *Application of Biomarkers in Cancer Epidemiology* (IARC Scientific Publications no. 142). Lyon: International Agency for Research on Cancer, 1997.
- Truett J, Cornfield J, Kannel W. A multivariate analysis of the risk of coronary heart disease in Framingham. *J Chron Dis* 1967; 20: 511-24.
- Blot WJ, McLaughlin JK, Devesa SS, Fraumeni JF Jr. Cancers of the oral cavity and pharynx. In: Schottenfeld D, Fraumeni JF Jr, eds. *Cancer Epidemiology and Prevention*. New York: Oxford University Press, 1996: pp. 666-80.
- Tucker JD, Eastmond DA et al. Cytogenetic end-points as biological dosimeters and predictors of risk in epidemiological studies. In: Toniolo P, Boffetta P, Shuker DEG, Rothman N, Hulka B, Pearce N, eds. *Application of Biomarkers in Cancer Epidemiology* (IARC Scientific Publications no. 142). Lyon: International Agency for Research on Cancer, 1997: pp. 185-200.
- Bonassi S, Hagmar L, Stromberg U et al. Chromosomal aberrations in lymphocytes predict human cancer independently of exposure to carcinogens. *Cancer Res* 2000; 60: 1619-25.
- Ross RK, Yuan J-M, Yu MC et al. Urinary aflatoxin biomarkers and risk of hepatocellular carcinoma. *Lancet* 1992; 339: 943-6.
- Rothman KJ, Greenland S. *Modern Epidemiology*. 2nd edn. Philadelphia: Lippincott-Raven Publishers, 1996.
- Smith G, Wolf CR. Cytochrome O430 CYP2D6. In: Malats N, Lang M, d'errico A, Caporaso N, Cuzick J, Boffetta P, eds. *Metabolic Polymorphisms and Susceptibility to Cancer* (IARC Scientific Publications no. 148). Lyon: International Agency for Research on Cancer, 1999: pp. 209-29.
- Vincis P. Sources of variation in biomarkers. In: Shuker DEG, Rothman N, Hulka B, Pearce N, eds. *Application of Biomarkers in Cancer Epidemiology* (IARC Scientific Publications no. 142). Lyon: International Agency for Research on Cancer, 1997: pp. 59-71.
- Schulte PA, Perera FP. Transitional Studies. In: Boffetta P, Shuker DEG, Rothman N, Hulka B, Pearce N, eds. *Application of Biomarkers in Cancer Epidemiology* (IARC Scientific Publications no. 142). Lyon: International Agency for Research on Cancer, 1997: pp. 19-29.
- Sanzella RM, Weston A, Perera FP et al. Interlaboratory comparison of antisera and immunoassays for benzo[a]pyrene-diol-epoxide-1-modified DNA. *Carcinogenesis* 1988; 9: 1265-9.
- Nakachi K, Imai K, Hayashi S, Kawajiri K. Polymorphisms of the CYP1A1 and glutathione S-transferase genes associated with susceptibility to lung cancer in relation to cigarette dose in a Japanese population. *Cancer Res* 1993; 53: 2994-9.
- Vang Q, Kiboury MJ. Evolving methods in genetic epidemiology III. Gene-environment interaction in epidemiologic research. *Epidemiol Rev* 1997; 19: 33-43.
- Slattery ML, Potter JD, Samowitz W, Bigler J, Caan B, Leppert M, NAT2, GSTM-1, cigarette smoking, and risk of colon cancer. *Cancer Epidemiol Biomarkers Prev* 1998; 7: 1097-84.
- Talati E. International collaborative study on genetic susceptibility to environmental carcinogens. *Cancer Epidemiol Biomarkers Prev* 1999; 8: 727-8 (Letter).
- d'errico A, Malats N, Vincis P et al. Review of studies of selected metabolic polymorphisms and cancer. In: Vincis P, Malats N, Lang M, d'errico A, Caporaso N, Cuzick J, Boffetta P, eds. *Metabolic Polymorphisms and Susceptibility to Cancer* (IARC Scientific Publications no. 148). Lyon: International Agency for Research on Cancer, 1999: pp. 323-93.
- Ambrosone CB, Freudenheim J, Graham S et al. Cigarette smoking, N-acetyltransferase 2 genetic polymorphisms, and breast cancer risk. *JAMA* 1996; 276: 1494-501.
- Greenland S. Hierarchical regression for epidemiologic analyses of multiple exposures. *Environ Health Perspect* 1994; 102(Suppl. 8): 33-9.
- Perera FP, Weinstein IB. Molecular epidemiology and carcinogen-DNA adduct detection: new approaches to studies of human cancer causation. *J Chronic Dis* 1982; 35: 581-600.
- Vincis P, Malats N, Lang M et al. eds. *Metabolic Polymorphisms and Susceptibility to Cancer* (IARC Scientific Publications no. 148). Lyon: International Agency for Research on Cancer, 1999.

Received 11 October 2000; accepted 17 October 2000.

Correspondence: Paolo Boffetta, Chief, Unit of Environmental Cancer Epidemiology, International Agency for Research on Cancer, 150 cours Albert-Thomas, 69008 Lyon, France (fax: +33 4-72738575; e-mail: boffetta@iarc.fr).